

# AVALIAÇÃO DE USABILIDADE ESTENDENDO O SUS COM PERGUNTAS ABERTAS: um estudo de caso

*USABILITY ASSESSMENT EXTENDING SUS WITH OPEN QUESTIONS: a case study*

PORTO, Pedro Burgos Monteiro; Bacharel; Universidade Federal do Ceará (UFC)

peburgosux@gmail.com

LEITE JUNIOR, Antonio Jose Melo; Doutor; Universidade Federal do Ceará (UFC)

melojr@virtual.ufc.br

GOMES, George Allan Menezes; Doutor; Universidade Federal do Ceará (UFC)

george@virtual.ufc.br

CHICCA JUNIOR, Natal Anacleto; Doutor; Universidade Federal de Pernambuco (UFPE)

natal.chicca@ufpe.br

## Resumo

O *System Usability Scale* (SUS) permite uma visão geral da usabilidade do sistema de forma rápida e a um baixo custo. Porém, por tratar apenas de dados quantitativos, ele pode ser considerado um tanto limitado em relação a aspectos gerais de uso do software. O presente artigo expõe o uso do SUS, porém agora incluindo também perguntas abertas, para coletar os dados e executar as etapas necessárias para analisar as informações que foram obtidas. Como exemplo, a proposta foi aplicada para avaliar um software educacional, resultando em informações sobre a percepção do uso pelos usuários e quais os problemas de usabilidade que eles encontraram. Os resultados obtidos retratam uma visão mais geral da usabilidade do sistema bem como outros mais contextuais e específicos sobre determinadas situações vivenciadas pelos usuários.

**Palavras-Chave:** procedimentos, SUS, usabilidade de software e percepção do usuário.

## Abstract

*The System Usability Scale (SUS) allows an overview of the system's usability quickly and at a low cost. However, as it only deals with quantitative data, it can be considered limited to general aspects of using the software. This article presents the use of SUS (System Usability Scale), now including open questions, to collect data and perform the necessary steps to analyze the information obtained. As an example, the proposal was applied to evaluate educational software, resulting in information about users' perception of its use and what usability problems they encountered. The results obtained portray a more general view of the system's usability as well as more contextual and specific ones to certain situations experienced by users.*

**Keywords:** procedures, SUS, software usability and user perception.

## 1 Introdução

Os conhecimentos sobre usabilidade surgiram de diversas áreas. Nielsen (1994), por exemplo, cita as áreas do design de interfaces, da interação humano-computador e da ergonomia como campos que influenciaram o que atualmente se entende como usabilidade. Para Petrie e Bevan (2009) é necessário que o processo de desenvolvimento de sistemas se adapte para enfrentar as dificuldades de compreender completamente os requisitos dos usuários, e desenvolver sistemas que forneçam experiências usáveis e agradáveis para os usuários.

Apesar disso, melhorar a usabilidade de um software pode trazer benefícios. Nielsen (1994) cita diversos casos documentados em que melhorias na usabilidade de um sistema resultaram em uma redução dos gastos da empresa, além de também conseguir aumentar a satisfação de uso e a facilidade em aprender a utilizar um determinado software.

Segundo Boucinha e Tarouco (2013), para ganhar aceitação no mercado atual, não basta mais apenas satisfazer os requisitos técnicos e as funcionalidades para as quais um sistema foi concebido; os usuários esperam que ele seja fácil de usar e que possam aprender rapidamente a utilizá-lo. Conforme Thomas (1996) menciona, nem sempre é possível aplicar um teste de usabilidade completo dentro de um ambiente industrial devido a restrições de tempo e custos. Brooke (1996) aponta que, nesses casos, frequentemente, tudo o que é preciso é uma indicação geral do nível geral de usabilidade de um sistema comparado com os seus competidores ou seus predecessores.

Neste cenário, tornou-se importante a busca por métodos de avaliação de usabilidade cada vez mais rápidos e de baixo custo como, por exemplo, o *System Usability Scale* (SUS) de Brooke (1996). O SUS, segundo Martins *et al.* (2015), possui diversas características que fazem ser uma boa escolha para praticantes de usabilidade geral e citam como principais vantagens: (I) a facilidade para preencher o questionário; (II) a facilidade de pessoas de diversas áreas compreenderem os resultados; (III) uma ferramenta de baixo custo por não ser um questionário proprietário. No entanto, devido a sua natureza quantitativa, utilizando apenas perguntas fechadas, complementar o SUS com a inclusão de perguntas abertas pode ser útil não apenas devido à possibilidade de coletar *feedbacks* mais subjetivos como também possibilitar um contraste com as demais respostas quantitativas fornecidas pelo mesmo usuário.

A complementação de perguntas dentro de uma metodologia que possua em suas etapas a análise desses diferentes tipos de dados de forma simples, rápida e objetiva pode ser uma alternativa para ajudar equipes de desenvolvimento de software a obterem melhorias no seu software a partir do uso de avaliações de usabilidade com usuários. Desta forma, o presente artigo se propõe a demonstrar uma proposta de procedimentos de usabilidade com base no SUS de Brooke (1996) e na resposta do usuário a perguntas abertas.

## 2 Definição de usabilidade

A seguir serão apresentados estudos que embasam teoricamente a definição do termo usabilidade que será utilizado no artigo, trazendo também métodos de avaliação de usabilidade qualitativos e quantitativos tendo foco principal na utilização do SUS de Brooke (1996).

A usabilidade, de acordo com Nielsen (1994), é definida como sendo o quão bem usuários conseguem utilizar uma funcionalidade. Ele também afirma que a usabilidade se aplica a todos os aspectos de um sistema no qual um ser humano pode interagir, indo desde a instalação e manutenção até o uso propriamente dito de um sistema. Percebe-se, portanto, a visão abrangente

de Nielsen sobre a atuação da usabilidade na utilização de um recurso ou funcionalidade.

Da mesma maneira, Shneiderman (2000) argumenta pela proposição da usabilidade universal em que, entre todos os domicílios com serviços de comunicação e informação, mais de 90% de usuários são bem sucedidos ao utilizar esses serviços ao menos uma vez por semana. Para tentar alcançar tal objetivo, Shneiderman (2000) destaca principalmente a importância em analisar a variedade da tecnologia, diversidade dos usuários e as lacunas no conhecimento dos mesmos. Ele também ressalta que acomodar um espectro mais amplo de situações de uso força os pesquisadores a considerarem uma gama maior de designs e, frequentemente, leva a inovações que beneficiam todos os usuários.

Com base nas informações apresentadas, é possível afirmar que aperfeiçoamentos na usabilidade não são apenas um modo de melhorar a forma atual de uso com os usuários atuais, mas também uma possibilidade de ampliar e democratizar essa utilização em contextos mais diversos e variados, para além dos que já haviam sido originalmente pensados.

A NBR 9241-210 (Associação Brasileira de Normas Técnicas, 2011) considera que a usabilidade é a medida na qual um produto pode ser usado por determinados usuários para alcançar certos objetivos com eficácia, eficiência e satisfação em um contexto específico de uso. Como eficácia ela especifica que equivale a acurácia e a completude com as quais usuários alcançam objetivos específicos. No caso da eficiência, ela define como sendo recursos utilizados em relação aos resultados alcançados. Por fim, satisfação é descrita como a medida a qual as respostas físicas, cognitivas e emocionais que resultam do uso de um sistema, produto ou serviço pelo usuário atingem as necessidades e expectativas deste mesmo usuário.

Sendo assim, percebe-se que, para compreender efetivamente a usabilidade de algo, é necessário que também se entenda quais são os objetivos buscados nesse uso, o modo com que eles são alcançados e de que forma o contexto de uso atua impondo restrições e limitando o fenômeno a ser analisado.

Para Scholtz (2004), enquanto teoricamente qualquer produto de software pode ter sua usabilidade avaliada, essa avaliação provavelmente não irá produzir bons resultados a menos que sejam seguidos procedimentos de engenharia de usabilidade. Segundo o autor, a engenharia de usabilidade é a disciplina que fornece métodos estruturados para alcançar a usabilidade no design de interfaces do usuário durante o desenvolvimento de um produto, na qual a avaliação de usabilidade é parte desse processo. A respeito da avaliação de usabilidade, Nielsen (1994) afirma que buscar aspectos mensuráveis da usabilidade de um sistema, mesmo não utilizando completamente os aspectos formais de avaliação, é mais indicado do que buscar um sentimento caloroso e difuso de ser amigável ao usuário.

Considerando isso, pode-se afirmar que um dos principais pontos para a devida avaliação de usabilidade depende de como os aspectos mensuráveis foram definidos de forma clara, mesmo que não seguindo todos os formalismos existentes dentro da engenharia de usabilidade.

## **2.1. Avaliação de usabilidade**

A avaliação de usabilidade, segundo Nielsen (1994, p.26), inicia com a “definição de um conjunto representativo de tarefas-teste nas quais diferentes atributos de usabilidade podem ser medidos”. Sobre esses atributos, o autor lista cinco:

- Facilidade de aprendizagem, o quanto o sistema permite que o usuário aprenda rapidamente a utilizá-lo;
- Eficiência, uma vez aprendido o sistema, se ele permite uma alta produtividade;
- Memorabilidade, o quanto o sistema permite a um usuário casual utilizá-lo novamente após um tempo sem uso;
- Erros, que aborda a forma com que os usuários são interrompidos e conseguem se recuperar dos erros decorrentes do uso do sistema;
- Satisfação, sobre o quanto os usuários do sistema subjetivamente gostam de utilizá-lo.

Os atributos de usabilidade de um software, segundo Scholtz (2004), nem sempre devem ser considerados de forma igual, pois, dependendo do tipo de aplicação, um deles pode se tornar mais importante para o uso do que os demais. Petrie e Bevan (2009), assim como Scholtz (2004), afirmam que a usabilidade é relativa aos usuários, objetivos e contextos de uso que são apropriados para o conjunto particular de circunstâncias, e que os requisitos ou critérios acerca dos atributos da usabilidade vão variar fortemente de acordo com essas circunstâncias. Para Scholtz (2004), se um software não é utilizado de forma frequente, se torna essencial que os usuários possam lembrar as ações necessárias para realizar as tarefas desejadas. Caso o tempo seja crítico para a aplicação, então a eficiência se torna prioridade, assim como a prevenção de erros.

Assim sendo, os atributos da usabilidade permitem uma análise mais precisa e estruturada sobre os principais pontos que os usuários podem encontrar dificuldades ao utilizar o sistema dentro de seu contexto de uso. Além disso, entender e priorizar melhorias nos atributos mais relevantes ao tipo de software utilizado facilita a realização de avaliações de usabilidade mais objetivas e que trazem mais benefícios aos usuários.

## **2.2. Métodos de avaliação de usabilidade**

Considerando-se os atributos da usabilidade, torna-se também importante conhecer sobre quais são os métodos de pesquisa que podem avaliar tais atributos. Esses métodos podem ser divididos primeiramente em três categorias: as pesquisas qualitativas, as pesquisas quantitativas e as pesquisas de métodos mistos. Tais categorias são discutidas a seguir.

### **2.2.1. Pesquisas quantitativas**

A pesquisa quantitativa, segundo Cresswell (2014), é uma abordagem para testar teorias objetivas ao examinar as relações entre variáveis. Essas variáveis, por sua vez, podem ser medidas, tipicamente por meio de instrumentos, para que os dados numéricos possam ser analisados utilizando procedimentos estatísticos. O autor também afirma que esse tipo de pesquisa fornece uma descrição numérica de tendências, atitudes ou opiniões de uma população ao estudar uma amostra dessa população. A partir dos resultados da amostra, o pesquisador generaliza ou conclui inferências para a população.

A pesquisa quantitativa, a partir de tais definições, se caracteriza pela redução das relações existentes nas informações buscadas a um conjunto reduzido de variáveis, na qual se possibilita mensurar e avaliar tendências e opiniões através do estudo de uma pequena amostra da população total.

Acerca especificamente de pesquisas quantitativas sobre usabilidade, podem-se citar as

avaliações baseadas em modelos que permitem, segundo Petrie e Bevan (2009), prever medidas como o tempo para completar uma tarefa ou a dificuldade em aprender a utilizar a interface. Um dos modelos citados pelos autores é *Goals Operators Methods and Selections* (GOMS). Para Scholtz (2004), o GOMS é um modelo que busca simular e decompor as ações do usuário com base nos métodos que serão utilizados para ele chegar a seu objetivo, e como esses métodos são selecionados através das regras de seleção existentes.

De acordo com Scholtz (2004), o modelo GOMS deve ser utilizado apenas para avaliar a eficiência do aspecto procedural da usabilidade, mas não pode avaliar potenciais erros devido ao design da tela ou terminologia, pois segundo o autor esse método não consegue avaliar potenciais erros devidos ao design da tela ou terminologia, e que a preparação para utilizar métodos de avaliação por meio de modelos exige etapas demoradas. Petrie e Bevan (2009) também destacam que a avaliação de usabilidade através de modelos e simulações é mais indicada quando não é possível realizar testes com usuários e o tempo para completar as tarefas é algo crítico para o sistema.

Um dos principais métodos quantitativos para pesquisar sobre a usabilidade de um sistema é a utilização de questionários, cuja avaliação visa coletar a opinião do usuário sobre determinados pontos do uso do sistema. Nielsen (1994) destaca que as vantagens da utilização de questionários se relacionam ao seu baixo custo e à sua capacidade de coletar um grande volume de respostas com pouco gasto de tempo. O autor destaca a necessidade do questionário ser projetado para ser de fácil interpretação pelos usuários, além de também evitar possuir um grande número de perguntas para não ser irritante e obter uma baixa taxa de resposta.

Apesar de existirem outros modelos de questionários voltados para usabilidade, tais como o SUMI (1993) e o QUIS (1988), este artigo vai focar na utilização do SUS de Brooke (1996). A justificativa pela escolha do SUS é que por possuir apenas 10 perguntas (Tabela 1), se torna um método mais rápido, quando comparado aos questionários SUMI e QUIS, que necessitam respectivamente de 50 e 90 perguntas. Dessa forma, o SUS é um método de avaliação de usabilidade que está mais alinhado ao objetivo da pesquisa em propor uma metodologia de rápida aplicação e execução.

Tabela 1 – As 10 perguntas utilizadas no SUS

1. Eu acho que gostaria de usar esse sistema com frequência.
2. Eu acho o sistema desnecessariamente complexo.
3. Eu achei o sistema fácil de usar.
4. Eu acho que precisaria de ajuda de uma pessoa com conhecimentos técnicos para usar o sistema.
5. Eu acho que as várias funções do sistema estão muito bem integradas.
6. Eu acho que o sistema apresenta muita inconsistência.
7. Eu imagino que as pessoas aprenderão como usar esse sistema rapidamente.
8. Eu achei o sistema atrapalhado de usar.
9. Eu me senti confiante ao usar o sistema.
10. Eu precisei aprender várias coisas novas antes de conseguir usar o sistema.

Fonte: adaptado de Brooke (1996)

O SUS é uma questionário voltado a avaliar a usabilidade de um software, por meio de uma escala Likert de 5 pontos. De acordo com Dalmoro e Vieira (2013), o modelo de escala Likert (1932) se caracteriza pelo seu caráter bidimensional, que ao longo de cinco pontos vai de “aprovo fortemente” até “desaprovo fortemente”, com um ponto neutro no meio da escala conforme

representado pela Figura 1. Dalmoro e Vieira (2013) também destacam que a escala Likert original apenas considerava escala com 5 pontos, sendo escalas com outros números de opções de respostas consideradas “tipo Likert”.

Figura 1 – Escala Likert de 5 pontos



Fonte: Likert (1932, apud. Petrie; Bevan, 1996)

O SUS também permite calcular uma pontuação, o índice SUS, que vai de 0 a 100 com base nas respostas obtidas pelo usuário para se ter uma visão geral da usabilidade do sistema. Segundo Bangor, Kortum e Miller (2009), é possível fazer uma correspondência entre índice SUS obtido e determinados adjetivos. Tais correspondências são:

- Pior imaginável, quando o índice SUS for aproximadamente 12.5;
- Horrível, ao índice atingir aproximadamente 20.3;
- Pobre, com aproximadamente 35.7;
- Ok, com índice próximo a 50.9;
- Bom, quando o índice chegar a aproximadamente 71.4;
- Excelente, com resultado próximo de 85.5;
- Melhor imaginável, quando aproximadamente 90.9 e acima.

### 2.2.2. Pesquisas qualitativas

A pesquisa qualitativa, segundo Cresswell (2014), é uma abordagem para explorar e entender o significado que indivíduos ou grupos atribuem a um problema social ou humano. De acordo com o autor, essa pesquisa envolve análise de dados construídos indutivamente de temas particulares para temas gerais, e o pesquisador fazendo interpretações sobre o significado dos dados. Para Cresswell (2014) as características básicas em pesquisa qualitativa são:

- Ambientação natural, em que a coleta de informações acontece dentro do contexto das pessoas e não em laboratórios;
- Pesquisador como instrumento chave, pois mesmo utilizando outros instrumentos para coleta de dados, o próprio pesquisador acaba sendo a principal fonte da coleta;
- Múltiplas fontes de dados, já que ao invés de depender de uma única fonte de dados, essas pesquisas coletam múltiplas fontes diferentes, tais como entrevistas, observações, documentos e informações audiovisuais;
- Análise de dados indutiva e dedutiva, os pesquisadores qualitativos utilizam processos indutivos e dedutivos para trabalhar com os temas percebidos e dados obtidos até durante a pesquisa- de forma a encontrar as informações que faltam;

- Significado dos participantes, durante a pesquisa o foco permanece em entender o significado percebido pelos participantes, e não o significado dos pesquisadores;
- Design emergente, indicando que alguns processos nesse tipo de pesquisa podem sofrer mudanças conforme o andamento da pesquisa em campo;
- Reflexividade, pois o próprio pesquisador questiona o quanto as suas circunstâncias pessoais podem moldar a direção do estudo;
- Relato holístico, ao relatar múltiplas perspectivas e os diversos fatores envolvidos na situação para ajudar a desenvolver um quadro mais completo sobre o problema investigado.

Sendo assim, de acordo com Nielsen (1994), um dos métodos de pesquisa qualitativa que pode ser utilizado para avaliar a usabilidade e satisfação do uso de um sistema é a entrevista com usuários. A entrevista é definida como um método exploratório e que, devido a suas perguntas abertas, permite aos usuários explicarem seus pontos de vista de forma mais detalhada. Nielsen (1994) também afirma que ela pode ser realizada por telefone, apesar de normalmente ser executada presencialmente no local do usuário. Para o autor, diferentemente de questionários, as entrevistas possuem mais flexibilidade para o entrevistador explicar e reformular perguntas difíceis, apesar dos questionários permitirem que se obtenha um número muito maior de respostas dos usuários geralmente gastando menos tempo.

Outro exemplo de pesquisa qualitativa em usabilidade é a inspeção heurística. Nielsen (1994) define a inspeção heurística como uma inspeção sistemática para o design de interface de usuário para usabilidade que busca encontrar problemas a serem resolvidos como parte de um processo de design iterativo. A inspeção heurística envolve um pequeno conjunto de avaliadores examinando a interface e julgando sua conformidade com princípios de usabilidade reconhecidos (as heurísticas).

Petrie e Bevan (2009) afirmam que a avaliação heurística, assim como originalmente foi proposta por Nielsen (1994), vem sendo adaptada de diferentes formas, já que ao invés de inspecionar elementos individuais ela é frequentemente realizada ao solicitar ao avaliador ir passo a passo através de tarefas comuns aos usuários.

Dessa forma, podemos considerar que a inspeção heurística é um tipo de pesquisa qualitativa que se vale do conhecimento especializado tanto para avaliar a interface quanto para a escolha e construção das heurísticas que serão utilizadas durante a inspeção. Isso permite que esse tipo de pesquisa traga uma perspectiva especializada para analisar a usabilidade do sistema.

Testar diretamente o sistema com usuários e as tarefas que serão realizadas é outro exemplo de pesquisa qualitativa em usabilidade. Segundo Petrie e Bevan (2009), em avaliações com usuários, eles são orientados a realizar tarefas nas quais o sistema é projetado para suportar em situações realistas, ou mais realísticas quanto possível. De acordo com os autores, avaliações com usuários podem ser utilizadas tanto para compreender o comportamento e intenções dos usuários quanto para medir a usabilidade ou acessibilidade do sistema.

Com base nas informações apresentadas, percebe-se a flexibilidade sobre quais informações podem ser obtidas ao realizar avaliações com usuários realizando tarefas reais pelo sistema. Isso se deve principalmente ao sistema estar sendo avaliado o mais próximo do contexto real de uso do que em outros testes, como a inspeção heurística, por exemplo, e permitir assim uma coleta de dados mais rica em informações contextuais.

Portanto, conclui-se que a pesquisa qualitativa é uma categoria de pesquisa cuja riqueza dos resultados obtidos provêm em se obter não apenas informações relevantes aos problemas da pesquisa, mas toda a contextualização necessária para se coletar essas informações de uma forma mais completa quanto o possível dentro das circunstâncias do problema.

### **2.2.3. Pesquisas de métodos mistos**

A utilização de perguntas abertas, mesmo quando aplicadas em um questionário, pode ser uma forma de obtenção de dados mais qualitativos do que seria obtido com apenas perguntas fechadas. Para Harland e Holey (2011), a inclusão de perguntas abertas dentro de questionários quantitativos pode adicionar riqueza aos dados, melhorar, e talvez possivelmente mudar a interpretação das descobertas. Os autores também destacam que uma possível vantagem de uma pergunta aberta escrita autoadministrada, quando comparado com uma entrevista cara a cara, é a possibilidade de eliminar uma grande porção do viés.

A pesquisa de métodos mistos, conforme a definição de Cresswell (2014), é uma abordagem de investigação envolvendo coletar tanto dados quantitativos quanto dados qualitativos. Nesse tipo de pesquisa se procura obter um entendimento mais completo do problema de pesquisa, combinando as abordagens qualitativas e quantitativas, em uma tentativa de buscar compensar as fraquezas que cada uma possui quando aplicada de forma separada. Os dados qualitativos podem ajudar a esclarecer determinados pontos do contexto de uma pesquisa quantitativa. O autor cita um caso em que a utilização de métodos mistos produziu resultados convergentes mesmo com a utilização de diferentes métodos, enquanto em outras situações os métodos produziram resultados divergentes e devido a isso possibilitou ao pesquisador a oportunidade de produzir uma explicação mais elaborada sobre o motivo dessas divergências.

Cresswell (2014) destaca três modelos principais de pesquisas de métodos mistos: os paralelos convergentes, os explanatórios sequenciais e os exploratórios sequenciais:

- Nos métodos mistos paralelos convergentes o pesquisador converge ou mistura métodos quantitativos e qualitativos, coletando ambos os tipos de dados simultaneamente para permitir posteriormente uma interpretação dos resultados obtidos;
- Nos métodos mistos explanatórios sequenciais o pesquisador conduz primeiramente uma pesquisa quantitativa, analisa os resultados e a partir disso constrói em cima dos resultados para explicá-los com mais detalhes através da pesquisa qualitativa. A nomenclatura desses métodos se deve ao fato que são considerados explanatórios devido aos resultados iniciais qualitativos serem explicados pelos dados qualitativos, e sequencial devido a uma etapa de pesquisa ocorrer posteriormente ao que foi obtido como resultado anteriormente;
- Nos métodos mistos sequenciais exploratórios, primeiro ocorre uma etapa qualitativa para ser seguida de uma etapa quantitativa. Eles permitem explorar inicialmente as visões dos participantes para, com base na análise dos dados qualitativos obtidos, construir mais satisfatoriamente a coleta dos dados quantitativos.

As pesquisas de métodos mistos surgem como uma possibilidade de coletar dados qualitativos e quantitativos da forma que melhor se adapte ao contexto e restrições do problema de pesquisa. A partir disso, presume-se que o pesquisador consiga ter uma visão mais completa do que teria ao aplicar apenas pesquisas qualitativas ou apenas pesquisas quantitativas, já que cada uma dessas permite uma visão diferente sobre a situação analisada.



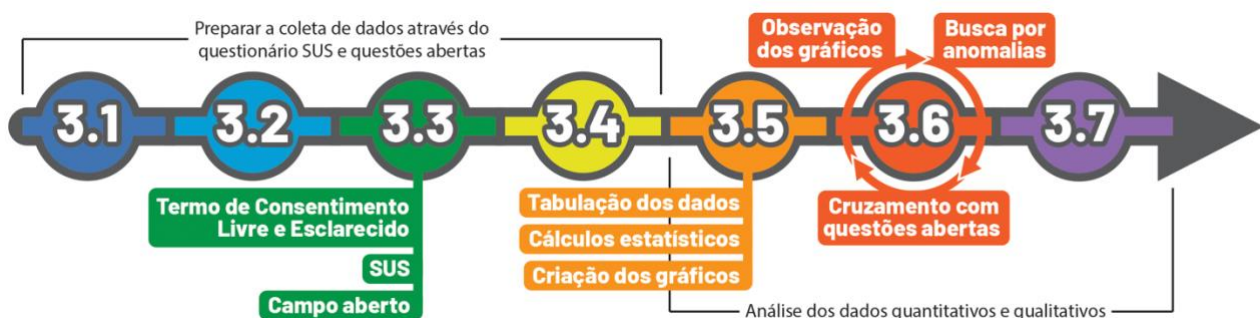
### 3 Proposta de procedimentos de avaliação de usabilidade de software

O procedimentos propostos foram desenvolvidos a partir de um levantamento bibliográfico e seleção de autores para cada uma das etapas elaboradas. A proposta utiliza o SUS através de abordagens quantitativa e qualitativa. Para isso, aplica um método misto de avaliação explanatório sequencial, baseado em Cresswell (2014), no qual o pesquisador conduz primeiramente uma pesquisa quantitativa, analisa os dados e, posteriormente, constrói em cima dos resultados para explicá-los com mais detalhes através de uma pesquisa qualitativa.

A proposta foi elaborada utilizando, como base principal, três das características essenciais da pesquisa qualitativa de Cresswell (2014): ambientação natural, em que a coleta de informações acontece dentro do contexto das pessoas e não em laboratórios; análise de dados indutiva e dedutiva; e foco no significado percebido pelos participantes – e não pelos pesquisadores. Em complemento, dado o caráter prático eminentemente da avaliação de usabilidade, foi adotado o modelo GOMS de Scholtz (2004), que busca simular e decompor as ações do usuário com base nos métodos que serão utilizados para ele chegar a seu objetivo, e como esses métodos são selecionados através das regras de seleção existentes.

Os procedimentos propostos consistem em sete etapas: (3.1) estabelecimento do perfil; (3.2) sensibilização dos usuários do perfil; (3.3) criação do questionário; (3.4) disponibilização do questionário e acesso pelos usuários/perfis; (3.5) análise qualitativa; (3.6) avaliação quantitativa; (3.7) geração de resultados e observações. Enquanto as primeiras quatro etapas buscam preparar a coleta de dados através do questionário SUS e questões abertas, as últimas três etapas visam a análise dos dados quantitativos e qualitativos obtidos, buscando obter uma interpretação mais rica através da mesclagem dos dois métodos. A etapa de avaliação, por sua vez, ocorre de forma cíclica devido ao seu caráter iterativo, em que os resultados obtidos são utilizados em um constante processo de análise, que continua se repetindo dentro das limitações de tempo estipuladas para a execução da proposta. O fluxo das etapas da proposta de avaliação de usabilidade de software está representado na Figura 2.

Figura 2 – Etapas da proposta de avaliação de usabilidade de software



Fonte: Elaborado pelos autores

Cada uma das etapas serão detalhadas a seguir:

#### 3.1. Estabelecimento do perfil

A primeira etapa estabelece qual é o perfil de usuário mais adequado aos objetivos da pesquisa. O perfil, baseado em Nielsen (1994), deve ser construído para garantir que os

participantes selecionados sejam aqueles que já utilizaram o sistema ou que possuam condições mais próximas de uso quanto possível.

### **3.2. Sensibilização de usuários por perfil**

Os usuários selecionados são expostos ao sistema, utilizando-o para realizar tarefas comuns ao seu dia a dia. A etapa de sensibilização tem como objetivo preparar os usuários para responderem o questionário o qual será aplicado posteriormente.

### **3.3. Criação do questionário**

O questionário é um conjunto de perguntas, por meio impresso ou eletrônico, em que são coletados dados relacionados ao problema pesquisado. Ele deve conter preferencialmente três seções bem definidas:

#### **3.3.1 Termo de Consentimento Livre e Esclarecido**

O Termo de Consentimento Livre e Esclarecido visa informar o usuário sobre a forma com que os dados coletados serão utilizados ao longo da pesquisa em curso. O termo deve buscar obter o consentimento e esclarecer o usuário sobre a coleta e utilização das respostas fornecidas no questionário.

#### **3.3.2 System Usability Scale (SUS)**

No questionário são adicionadas dez perguntas relacionadas ao SUS, em que cada uma delas possui uma escala Likert de 5 pontos, indo de concordo totalmente a discordo totalmente, levando em consideração o trabalho original de Brooke (1996).

#### **3.3.3 Questões abertas**

O uso de questões abertas, respondidas de forma textual e subjetiva, conforme apontado por Harland e Holey (2011), busca ajudar o pesquisador a obter uma visão mais completa sobre as informações quantitativas que também foram obtidas durante o questionário.

### **3.4. Disponibilização do questionário e acesso pelos usuários/ perfis**

O questionário é disponibilizado aos usuários para que eles possam responder sobre suas percepções no uso do sistema. Durante o seu preenchimento, eles devem marcar suas respostas nas escalas das perguntas objetivas e descreverem textualmente nas perguntas abertas.

### **3.5. Análise Quantitativa**

Os dados quantitativos são tratados, tabulados para simplificar a sua manipulação e, a partir disso, são realizados cálculos estatísticos simples para indicar possíveis tendências nas respostas. Por fim, criam-se gráficos para facilitar a observação de padrões existentes.

### **3.6. Ciclo de avaliação**

Essa etapa ocorre de forma cíclica, em que o pesquisador observa os gráficos, buscando por anomalias que fujam dos padrões observados até então. Assim como também cruza os dados quantitativos com o que foi obtido no campo aberto do questionário. Sendo assim o ciclo engloba três etapas:

#### **3.6.1 Observação dos gráficos**

Verificar se os gráficos estão de acordo (ou não) com o esperado idealmente se torna um método bastante rápido de diagnosticar anomalias. Devido à alternância nas perguntas ímpares e pares durante o cálculo do SUS, o gráfico de barras criado a partir das perguntas ímpares do SUS vai

se aproximar mais do concordo totalmente, obtendo assim um padrão crescente no gráfico de barras; e os que são gerados a partir das perguntas pares irão se aproximar de um padrão decrescente.

### 3.6.2 Busca por anomalias

As anomalias são pontos que fogem do esperado, surgindo ao longo do ciclo de avaliação. Cabe ao pesquisador cruzar as diferentes fontes de informações obtidas que explique a ocorrência da anomalia detectada.

### 3.6.3 Cruzamento com questões abertas

Ao cruzar os dados obtidos através das questões abertas, o pesquisador consegue compreender a situação de forma mais completa. Ou seja, esse passo busca auxiliar o pesquisador durante o ciclo de avaliação na análise da anomalia detectada.

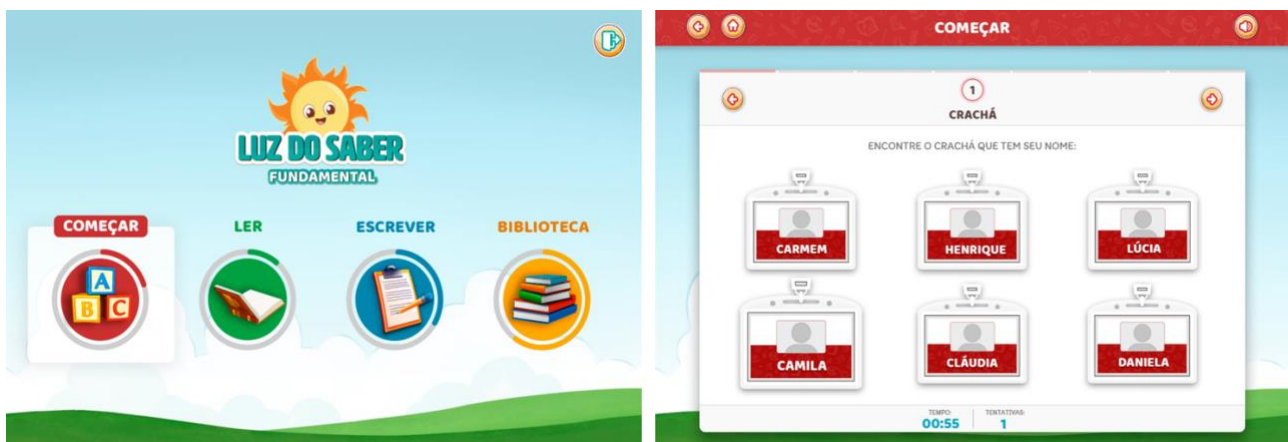
### 3.7. Geração de resultados e observações

Ao longo de todas as etapas da pesquisa, o pesquisador deve ter a constante preocupação em documentar todos os resultados que vão sendo obtidos com suas devidas observações. Isso possibilita ter mais detalhes acerca do contexto das informações obtidas, além de facilitar o gerenciamento do progresso da avaliação geral.

## 4 Estudo de caso: Luz do saber

O Luz do Saber (<https://luzdosaber.seduc.ce.gov.br/>) é um software educacional infantil que visa alfabetizar e letrar tanto crianças como jovens e adultos, empregando exercícios diversos baseados em gamificação (Figura 3). A versão atualizada estava prevista para uso através de um projeto-piloto no início de 2020, em algumas escolas públicas pré-selecionadas, com duração de seis meses para eventuais correções de problemas e, posteriormente, a disponibilização junto à comunidade em geral. No entanto, a pandemia de COVID-19 antecipou a liberação do mesmo a todos, através do acesso por smartphones.

Figura 3 – Capturas de telas da interface do Luz do Saber



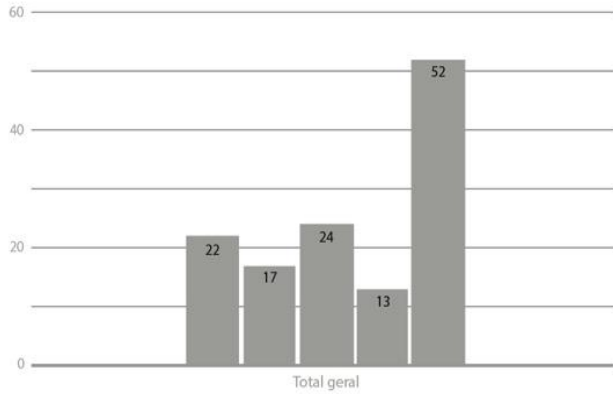
Fonte: Elaborado pelos autores

O Laboratório de Mídias Educacionais, vinculado à Universidade Federal do Ceará (UFC), junto ao Governo do Estado do Ceará, foram os responsáveis em aferir quais foram os impactos no uso e quais eram as dificuldades enfrentadas pelos usuários durante a utilização. Com isso, durante entre os meses de outubro e novembro de 2020 foi realizada uma pesquisa sobre a usabilidade do sistema junto a seus usuários. Para tanto, foi seguida especificamente os procedimentos apresentados na seção 3 e a execução de suas etapas serão descritas a seguir:

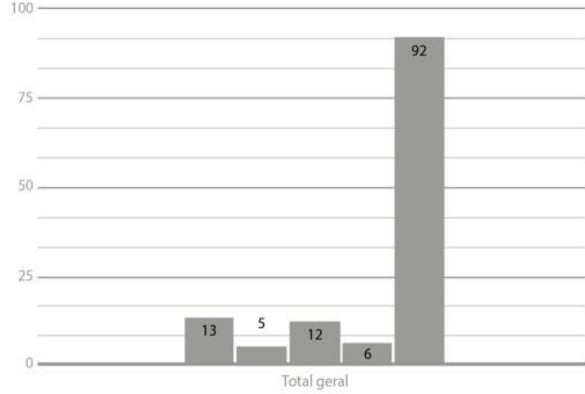
- **Estabelecimento do Perfil:** Inicialmente, levantando quais poderiam ser os diferentes usuários do software e quem poderia auxiliar durante a resposta aos questionários, foram estabelecidos quatro perfis de participantes possíveis: perfil A: criança usuária do Luz do Saber respondendo; perfil B: pai/ mãe/ responsável por criança usuária do Luz do Saber respondendo; perfil C: pai/ mãe/ responsável ajudando criança usuária do Luz do Saber a responder; perfil D: professor(a) respondendo.
- **Sensibilização dos usuários por perfil:** Devido ao software Luz do Saber Infantil já estar sendo utilizado, os responsáveis do projeto possuíam um meio de contatar os usuários ativos, precisando apenas divulgar o questionário para ser respondido, através dos canais de comunicação mantidos via Instagram e WhatsApp, por exemplo.
- **Criação do questionário:** O questionário foi criado de forma eletrônica, através da plataforma *Google Forms*, com treze perguntas sendo onze fechadas e duas abertas. Ele foi estruturado da seguinte forma: Termo de Consentimento Livre e Esclarecido; pergunta aberta para coletar o nome do usuário (opcional); pergunta fechada para separar por perfil de usuário (opcional); as dez perguntas fechadas do SUS; pergunta aberta sobre críticas, sugestões e elogios no final do questionário.
- **Disponibilização do questionário e acesso pelos usuários/ perfis:** O questionário foi disponibilizado para usuários do Luz do Saber através de um link no portal próprio do Luz do Saber, que direcionava os participantes para uma página no *Google Forms* em que os participantes poderiam responder durante o período de 20 de outubro até 20 de novembro de 2020. As respostas foram automaticamente armazenadas em uma planilha no *Google Sheets*, obtendo um total de 130 respostas ao final da coleta.
- **Análise Quantitativa:** Durante essa etapa foram calculados os índices SUS para cada resposta obtida, e com base nesses cálculos, foi extraído as médias aritméticas e desvio padrão para cada um dos perfis de usuário. Além disso, as médias foram interpretadas de acordo com a escala de Bangor *et al.* (2009).
- **Ciclo de avaliação:** Também foram criados gráficos relativos a quais foram as opções mais selecionadas para cada uma das perguntas utilizadas no SUS. Os resultados obtidos foram comparados ao que seria esperado idealmente no SUS, ou seja, crescente nas perguntas ímpares e decrescente nas perguntas pares. A partir desta análise, percebeu-se uma anomalia apresentada nos gráficos das perguntas 2, 4, 6 e 9 (conforme apresentado no Gráfico 1), pois tanto nos casos das perguntas pares quanto das perguntas ímpares, a opção 3 se destacou a ponto de impedir a ocorrência dos padrões esperados.

Gráfico 1 - Respostas obtidas nas perguntas 2, 4, 6 e 9 dos questionários avaliando o Luz do Saber

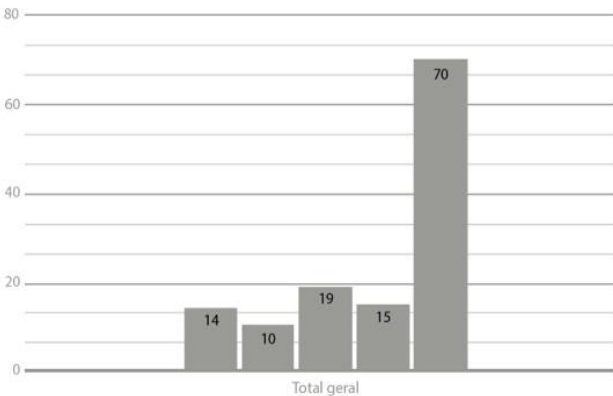
2. Considerei o Luz do Saber mais complexo do que o necessário (respostas de 1 a 5, da esquerda para direita).



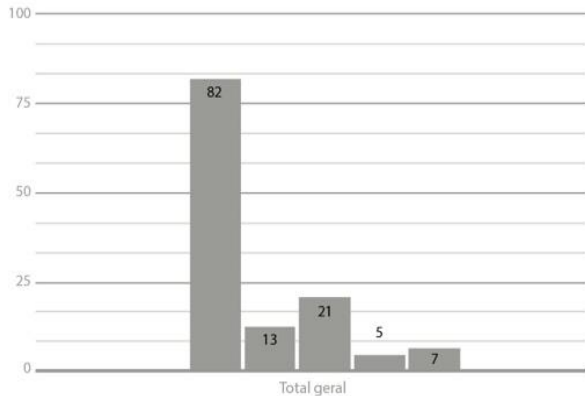
4. Acho que precisaria da ajuda de um técnico para conseguir utilizar o Luz do Saber (respostas de 1 a 5, da esquerda para direita).



6. Achei que o Luz do Saber tinha muitas inconsistências (respostas de 1 a 5, da esquerda para direita).



9. Senti-me muito confiante para utilizar o Luz do Solar (respostas de 1 a 5, da esquerda a direita).



Fonte: Elaborado pelos autores

Durante a análise das perguntas abertas, foi considerada a possibilidade de que usuários poderiam ter enviado respostas repetidas. Essa repetição foi detectada principalmente ao comparar as perguntas abertas no questionário, em que dois participantes preencheram três vezes de forma idêntica os campos de nome, perfil, questionário SUS e comentários. Considerando que esse fenômeno altera a composição dos dados quantitativos e dos gráficos gerados, foi necessário recalcular essas informações, conforme mostra a Tabela 2.

Tabela 2 - Médias de índice SUS geral e por perfil de usuário obtidas

Classificação segundo adjetivos de Bangor <i>et al.</i> (2009)	Média Índice SUS	Desvio Padrão Índice SUS
Média Geral considerando todos os perfis	75,1171875	23,64821616
<b>Separado por perfil de usuário</b>	<b>Média Índice SUS</b>	<b>Desvio Padrão Índice SUS</b>
Perfil A	86,66666667	11,27312438
Perfil B	69,73684211	22,33641338
Perfil C	72,33333333	25,24819329
Perfil D	79,56140351	23,51444658

Fonte: Elaborado pelos autores

A remoção das duplicatas resultou em alterações mínimas dos gráficos, portanto foi preciso investigar mais a fundo os dados quantitativos para descobrir quais eram os perfis de usuários que mais influenciaram para essa acentuação nas respostas em opções de igual distância entre concordo plenamente e discordo completamente. De acordo com a escala Likert, essa opção é considerada como “indeciso”, pois o usuário não consegue formar sua opinião sobre o assunto da pergunta. A Tabela 3 indica quantas respostas de cada perfil selecionou a opção indeciso.

Tabela 3 - Quantidade de respostas “indeciso” por cada perfil

Porcentagens de respostas “indeciso” por perfil	Pergunta 2 (24)	Pergunta 4 (12)	Pergunta 6 (16)	Pergunta 9 (19)	Pergunta 10 (17)
Perfil A (3 respostas)	1	0	1	1	1
Perfil B (30 respostas)	6	1	0	5	3
Perfil C (40 respostas)	11	8	10	8	6
Perfil D (57 respostas)	6	3	5	6	7

Fonte: Elaborado pelos autores

## 5 Resultados e discussão

O uso do índice SUS permitiu diagnosticar a existência de pontos mais críticos a serem melhorados. Além disso, segundo Brooke (2013), devido à confiabilidade do SUS, o software pode ser novamente avaliado após um determinado período. Os resultados obtidos anteriormente podem ser comparados para compreender se as mudanças realizadas realmente atingiram o objetivo de melhorar a usabilidade do Luz do Saber. No caso desta pesquisa foi utilizada a escala de adjetivos de Bangor, Kortum e Miller (2009) para qualificar os índices encontrados.

As etapas de análise quantitativa e os ciclos de avaliação foram fundamentais para alcançar os objetivos da pesquisa. Devido à tabulação e visualização dos gráficos em barras, por exemplo, foi possível perceber em quais perguntas os participantes tiveram maior dificuldade e quais eram as situações que fugiam do esperado, causando uma anomalia. A detecção de anomalias apresenta uma rica oportunidade para desenvolver melhorias e fornecer uma usabilidade aprimorada aos usuários.

Durante os ciclos de avaliação, a presença das questões abertas e fechadas trouxe mais possibilidades de análise. Elas permitiram ter um pouco mais de contexto sobre os dados quantitativos, obtidos nas outras perguntas, proporcionando uma oportunidade a mais para buscar explicações para as situações que fugiam do perfil ideal esperado e entender melhor sobre os índices SUS abaixo da média. Exemplificando, foi percebido que algumas pessoas cujas respostas resultaram em índices SUS baixos citaram no campo aberto alguns bugs ou dificuldades ao tentar acessar o Luz do Saber, o que pode ter influenciado a sua experiência de uso e conseqüentemente sua percepção sobre a usabilidade do Luz do Saber.

Além disso, a alta taxa de respostas “indeciso” pelo perfil C (responsável ajudando a criança a responder) trouxe a oportunidade de pesquisar e avaliar qual seria o motivo dessa alta taxa de

respostas. Um motivo possível seria uma dificuldade de compreensão da pergunta pelo perfil, mas também pode ser válido testar futuramente de que forma alterações e melhorias nos aspectos mencionados por essas perguntas, observando como as mudanças podem se refletir nas taxas de respostas do público, avaliando novamente o perfil C, em particular.

Da mesma forma, além de auxiliar nos ciclos de avaliação, a questão aberta, com sugestões e comentários para o Luz do Saber, permitiu coletar *feedbacks* que podem levar a melhorias não apenas no software, mas também aos processos educacionais em que ele é utilizado. Dentre as 64 respostas obtidas nessa questão, aproximadamente 18% delas se referiam ao material didático fornecido, constituído por livros normalmente utilizados em escolas públicas do Estado do Ceará. Também se destacaram comentários solicitando melhorias na acessibilidade, que apesar de terem sido apenas 3% dos respondentes, podem ser muito úteis para facilitar ainda a ampliação do uso do Luz do Saber. Por fim, 12% das respostas abordaram as atividades presentes no Luz do Saber e propuseram possíveis mudanças e melhorias: seja no aumento da variedade de modelos de atividades; maior autonomia do professor para a produção de mais atividades; e de tornar a gamificação mais fácil para as crianças. As sugestões identificadas são oportunidades para possíveis melhorias futuras.

Percebe-se, assim, que a análise qualitativa, baseada na avaliação dos campos abertos, consegue complementar a análise quantitativa, original do SUS. Assim, é possível enriquecer o modo com a usabilidade pode ser trabalhada nos sistemas e, ainda, propor melhorias para avaliação futura, considerando o público em geral ou perfis particulares.

## 6 Considerações finais

O artigo buscou apresentar e exemplificar uma proposta de procedimentos de avaliação de usabilidade de software, unindo a confiabilidade de dados quantitativos, obtidos por meio da utilização do SUS de Brooke (1996), e a contextualização fornecida pelos dados qualitativos através das perguntas abertas. Com base na utilização dos procedimentos propostos, conseguiu-se avaliar o Luz do Saber tanto com dados que retratam uma visão geral da usabilidade do sistema quanto outros mais contextuais e específicos sobre determinadas situações.

Como trabalhos futuros, vislumbra-se a possibilidade de validar os procedimentos a partir de outros exemplos de softwares, buscando aumentar a validade e a confiabilidade da proposta. Além disso, uma forma de enriquecer e tornar a proposta mais robusta será a adição de outras etapas, como a da análise dos dados ou alterações na forma de coleta dos dados qualitativos, apesar de que os objetivos da pesquisa são voltados à sua agilidade de aplicação e avaliação.

As contribuições esperadas por esse artigo têm como objetivo propor uma solução simples e rápida que agilize o processo de avaliação de usabilidade, obtendo, de forma rápida e confiável, dados quantitativos e qualitativos. Essa mesclagem de abordagens consegue, com baixo custo e em pouco tempo, obter uma visão mais ampla da usabilidade e da satisfação dos usuários sobre o software avaliado.

## Referências

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 9241-11: **Requisitos Ergonômicos para Trabalho de Escritórios com Computadores - parte 11- orientações sobre Usabilidade**. Rio de Janeiro: ABNT, 2002.

- BANGOR, A.; KORTUM, P.; MILLER, J. Determining what individual SUS scores mean: Adding an adjective rating scale. In: **Journal of usability studies**, v. 4, n. 3, 2009.
- BOUCINHA, R. M.; TAROUCO, L. M. R. Avaliação de Ambiente Virtual de Aprendizagem com o uso do SUS - System Usability Scale. In: **Revista Novas Tecnologias na Educação**, Porto Alegre, v. 11, n. 3, 2013.
- BROOKE, J. SUS – A quick and dirty usability scale. In: **Jordan, P.W. et al., Usability Evaluation in Industry**, pp. 189-94. London, UK: Taylor & Francis, 1996.
- BROOKE, J. SUS: a retrospective. In: **Journal of usability studies**, v. 8, p. 29-40, 2013.
- CRESWELL, John W. *et al.* **Research Design: Qualitative, quantitative, and mixed methods approaches**. 4.ed, 2014.
- DALMORO, M.; VIEIRA, K. M. Dilemas na construção de escalas Tipo Likert: o número de itens e a disposição influenciam nos resultados? In: **Revista gestão organizacional**, v. 6, n. 3, 2013.
- HARLAND, N.; HOLEY, E. Including open-ended questions in quantitative questionnaires—theory and practice. In: **International Journal of Therapy and Rehabilitation**, v. 18, n. 9, 2011.
- MARTINS, A. I.; ROSA, A. F.; QUEIRÓS, A.; ROCHA, N. P. European portuguese validation of the system usability scale (SUS). In: **Procedia Computer Science**, v. 67, p. 293-300, 2015.
- NIELSEN, J. **Usability engineering**. San Francisco, CA: Morgan Kaufmann, 1994.
- PETRIE, H.; BEVAN, N. **The Evaluation of Accessibility, Usability, and User Experience**. Boca Raton, USA: Crc Press, p. 20.1-20.14. 2009.
- SCHOLTZ, J. Usability evaluation. In: **National Institute of Standards and Technology**, v. 1, 2004.
- SHNEIDERMAN, B. Universal usability. In: **Communications of the ACM**, v. 43, n. 5, p. 84-91, 2000.
- THOMAS, B. Quick and dirty usability tests. In: **Usability evaluation in industry**, p. 107-114, London, UK: Taylor & Francis, 1996.